Contrôle Terminal – 07/05/2024 – 2h00
Statistiques – L1 de psychologie

Réponses

Veuillez rendre ce sujet et votre copie.

Numéro d'anonymat :

Le formulaire et la calculatrice sont autorisés. Merci d'indiquer dans la case Numéro d'anonymat*, ci-dessus, un numéro que vous reporterez aussi sur votre copie. Vous rendrez l'énoncé et votre copie, et pouvez soit répondre sur l'énoncé, soit détailler certaines questions sur la copie si vous avez besoin de plus de place. Le soin de la rédaction entrera en compte dans la notation mais dans les questions où des détails ne sont pas explicitement demandés, un résultat correct, donné sans détails de calcul sera accepté.*

Exercice 1 : Dépistage des troubles psychotiques

Les troubles psychotiques se dépistent lors d'un entretien avec un·e psychologue clinicien·ne. Toutefois, quand on cherche à diagnostiquer un grand nombre d'individus, on peut s'aider de questionnaires d'aide au dépistage, qui permettent de cibler les personnes auxquelles on fera passer l'entretien et de guider celui-ci.

Une large étude est menée en Europe pour établir un questionnaire de ce type, et vérifier si les réponses au questionnaire sont liées à la présence de troubles psychotiques. De nombreux·ses psychologues font ainsi passer un entretien de dépistage à un échantillon de 7 000 européen·ne·s, et leur font par ailleurs remplir le questionnaire, afin de comparer leur diagnostique avec les réponses au questionnaire ; on obtient ainsi les résultats ci-dessous :

Score	[40 ; 60[	[60 ; 80[	[80 ; 100[	[100 ; 120[	[120 ; 140[	[140 ; 160[
Effectif	22	617	2 891	2 834	606	30
Nombre de diagnostiques de troubles psychotiques	0	1	8	16	19	12

Par exemple, la troisième colonne de ce tableau indique que parmi les 7 000 personnes interrogées il y en a 2 891 qui ont eu un score entre 80 et 100 au questionnaire, et que parmi ces 2 891 personnes il y en a 8 qui ont été diagnostiqué·e·s comme souffrant de troubles psychotiques.

On supposera dans la suite que cet échantillon est représentatif de la population européenne.

1. On désigne par X le score obtenu à ce test et par Y la présence ou l'absence de troubles psychotiques. Ces variables sont-elles appariées ? *Justifiez brièvement votre réponse.*

Oui, elles sont appariées, car c'est pour les mêmes individus que l'on a mesurées ces deux variables.

2. Montrez (en indiquant les calculs effectués), qu'au sein de cet échantillon, les 56 personnes souffrant de troubles psychotiques ont un score moyen d'environ 121,79 au questionnaire, avec un écart type d'environ 20,62.

moyenne : $m(X) = \frac{\sum c_i n_i}{n} = \frac{50\times0+70\times1+90\times8+\cdots+150\times12}{56} = \frac{6\,820}{56} \simeq 121{,}79$

$m(X^2) = \frac{\sum c_i^2 n_i}{n} = \frac{50^2\times0+70^2\times1+90^2\times8+\cdots+150^2\times12}{56} = \frac{854\,400}{56}$

$Var(X) = m(X^2) - m(X)^2 = \frac{854\,400}{56} - \left(\frac{6\,820}{56}\right)^2 \simeq 425{,}38$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 20{,}62$

3. (a) Parmi les personnes de l'échantillon dont le **score** est **plus grand que 120**, quelle proportion souffre de troubles psychotiques ?

Il y a $606+30 = 636$ personnes dont le score est plus grand que 120, parmi lesquels $19+12 = 31$ ont été diagnostiqués comme souffrant de troubles psychotiques.
Cela correspond donc à la proportion $\frac{31}{636} \simeq 0{,}048\,7$.

(b) Justifiez que l'on peut utiliser la procédure du formulaire pour estimer (par intervalle de confiance) la proportion de personnes souffrant de troubles psychotiques, parmi l'ensemble des européen·ne·s dont le **score** est **plus grand que 120**. *Déterminer l'intervalle de confiance pour la confiance 95%.*

On a $n\,p_e(1-p_e) = 636 \times \frac{31}{636}\left(1-\frac{31}{636}\right) \simeq 29{,}489 > 10$, donc on peut utiliser la procédure du formulaire pour estimer la proportion de personnes psychotiques parmi l'ensemble des européen·ne·s qui ont plus que le score 120.
On a $F(1{,}96) \simeq 0{,}975$ d'où $z_\alpha \simeq 1{,}96$
d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}01674$ et $p_e - a_\alpha \simeq \frac{31}{636} - 0{,}016\,74 \simeq 0{,}032$ et $p_e + a_\alpha \simeq \frac{31}{636} + 0{,}016\,74 \simeq 0{,}065\,48$.
On estime donc que cette proportion est dans l'intervalle $[0{,}032; 0{,}06548]$ avec la confiance $c = 0{,}95$.

4. De même cet échantillon contient 6364 personnes dont le **score** est **plus petit que 120**, parmi lesquelles 25 souffrent de troubles psychotiques. En déduire une estimation (par intervalle de confiance) de la proportion de personnes souffrant de troubles psychotiques, parmi l'ensemble des européen·ne·s dont le score est plus petit que 120. *Déterminer l'intervalle de confiance pour la confiance 95%.*

On a $n\,p_e(1-p_e) = 6\,364 \times \frac{25}{6\,364}\left(1-\frac{25}{6\,364}\right) \simeq 24{,}902 > 10$, donc on peut utiliser la procédure du formulaire pour estimer la proportion de personnes psychotiques parmi l'ensemble des européen·ne·s qui ont moins que le score 120.
On a $F(1{,}96) \simeq 0{,}975$ d'où $z_\alpha \simeq 1{,}96$
d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}001537$ et $p_e - a_\alpha \simeq \frac{25}{6\,364} - 0{,}001\,537 \simeq 0{,}002\,391$ et $p_e + a_\alpha \simeq \frac{25}{6\,364} + 0{,}001\,537 \simeq 0{,}005\,465$.
On estime donc que cette proportion est dans l'intervalle $[0{,}002391; 0{,}005465]$ avec la confiance $c = 0{,}95$.

5. Pour chacune de ces affirmations, peut-on conclure, avec la confiance 95% qu'elle soit vraie, ou fausse ? Justifiez votre réponse en vous appuyant sur les intervalles de confiance calculés ci-avant. *En plus de cocher Vraie/Fausse/On-ne-peut-rien-affirmer, on vous demande de justifier votre réponse à partir des estimations des questions précédentes.*

Parmi les européen·ne·s qui ont un **score plus grand que 120**, la proportion qui souffrent de troubles psychotiques est plus grande que 4%.

☐Vraie ☐Fausse ☒ On ne peut rien affirmer

En effet, l'intervalle calculé en question 3b) contient à la fois des valeurs plus petites que 4% et des valeurs plus grandes que 4%.

Parmi les européen·ne·s qui ont un **score plus petit que 120**, la proportion qui souffrent de troubles psychotiques est plus grande que 4%.

☐Vraie ☒ Fausse ☐On ne peut rien affirmer

En effet, l'intervalle calculé en question 4 ne contient que des valeurs plus petites que 4%.

Parmi les européen·ne·s qui ont un **score plus grand que 120**, la proportion qui souffrent de troubles psychotiques est plus grande que parmi les européen·ne·s qui ont un **score plus petit que 120**.

☒ Vraie ☐Fausse ☐On ne peut rien affirmer

En effet, les intervalles des questions 3b et 4 ne se chevauchent pas : chaque valeur de l'intervalle de la question 3b est plus grande que chaque valeur de l'intervalle de la question 4.

Exercice 2 : Loi Normale Si les scores (notés X) obtenus à un questionnaire suivaient la loi $\mathcal{N}(100; 15)$, quels seraient les effectifs théoriques pour un échantillon de 7 000 individus ? *On demande de les indiquer dans le tableau ci-dessous*

Avec la calculatrice, on obtient $7\,000 \times \mathbb{P}[X < 60] \simeq 7\,000 \times 0{,}003\,830\,38 \simeq 26{,}813$,
$7\,000 \times \mathbb{P}[60 \leqslant X < 80] \simeq 7\,000 \times 0{,}087\,380\,8 \simeq 611{,}666$,
$7\,000 \times \mathbb{P}[80 \leqslant X < 100] \simeq 7\,000 \times 0{,}408\,789 = 2\,861{,}523$,
$7\,000 \times \mathbb{P}[100 \leqslant X < 120] \simeq 7\,000 \times 0{,}408\,789 = 2\,861{,}523$,
$7\,000 \times \mathbb{P}[120 \leqslant X < 140] \simeq 7\,000 \times 0{,}087\,380\,8 \simeq 611{,}666$
et $7\,000 \times \mathbb{P}[X \geqslant 140] \simeq 7\,000 \times 0{,}003\,830\,38 \simeq 26{,}813$. D'où les effectifs théoriques ci-dessous :

Score	[40 ; 60[	[60 ; 80[	[80 ; 100[	[100 ; 120[	[120 ; 140[	[140 ; 160[
Effectif théorique	26,813	611,666	2861,523	2861,523	611,666	26,813

Exercice 3 : Violence et troubles psychotiques

On s'intéresse dans cet exercice à l'idée répandue selon laquelle les personnes qui souffrent de troubles psychotiques seraient plus violentes que le reste de la population.

N'ayant pas les moyens d'effectuer des entretiens de diagnostique avec un grand nombre de personnes, les psychologues qui mènent cette étude choisissent d'utiliser un questionnaire comme celui de l'exercice 1, pour avoir une idée de la prévalence des troubles psychotiques au sein de différentes populations. On considère en effet que plus le score moyen à ce questionnaire est élevé au sein d'une population, plus cette population contient de personnes qui souffrent de troubles psychotiques.

1. On considère tout d'abord un groupe de 26 français·es majeur·e·s qui ont été condamné·e·s pénalement pour des actes de violence. En leur faisant passer le questionnaires, on obtient un score moyen de 103,61, avec un écart type de 19,83. *On supposera que parmi les personnes condamnées pour violence, ce score suit une loi normale.*

   En utilisant le formulaire qu'ils/elles avaient reçu lors de leurs études en L1 de psychologie, ces psychologues estiment, avec la confiance 95%, que le score moyen des français·es majeur·e·s condamné·e·s pour violence se trouve dans l'intervalle [95,44 ; 111,78].

   En présentant les calculs effectués, justifiez que c'est bien l'intervalle de confiance que l'on obtient à partir de cet échantillon.

   Comme le score est supposé suivre une loi normale, on peut utiliser la procédure du formulaire (estimation de moyenne pour un petit échantillon) :
   Comme $n = 26 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}025$ et $n - 1 = 25$ degrés de liberté (ddl)
   On lit $t_\alpha \simeq 2{,}0595$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 8{,}168$.
   On estime donc que le score moyen des personnes condamnées pour violence est dans l'intervalle $[103{,}61 - 8{,}168\,;\, 103{,}61 + 8{,}168] \simeq [95{,}44\,;\, 111{,}78]$ avec la confiance $c = 0{,}95$

2. À titre de comparaison on considère un groupe témoin de 200 français·es majeur·e·s. On les fait répondre au questionnaire et on obtient un score moyen de 98,35 avec un écart-type de 16,81.

   Déduisez-en un intervalle de confiance pour le score moyen de l'ensemble des majeur·e·s français·es, avec la confiance 95%.

   Comme n= 200 >30, on cherche z_α tel que $F(z_\alpha) = \frac{0{,}95+1}{2} = 0{,}975$
   On lit sur la table inverse que pour la confiance 0,95 on a $z_\alpha \simeq 1{,}96$ d'où $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 2{,}336$
   On estime donc que le score moyen des majeur·e·s est dans l'intervalle $[98{,}35 - 2{,}336\,;\, 98{,}35 + 2{,}336] \simeq [96{,}01\,;\, 100{,}69]$, avec la confiance $c = 0{,}95$.

3. Pour chacune de ces affirmations, peut-on conclure, avec la confiance 95% qu'elle soit vraie, ou fausse ? Justifiez votre réponse en vous appuyant sur les intervalles de confiance calculés ci-avant. *En plus de cocher Vraie/Fausse/On-ne-peut-rien-affirmer, on vous demande de justifier votre réponse à partir des estimations des questions précédentes.*

Le score moyen des français·es majeur·e·s condamnées pour violence est plus grand que 120.
☐Vraie ☒ Fausse ☐On ne peut rien affirmer
En effet, l'intervalle calculé en question 1 ne contient que des valeurs plus petites que 120.

Le score moyen des majeur·e·s français·es est plus petit que 120.
☒ Vraie ☐Fausse ☐On ne peut rien affirmer
En effet, l'intervalle calculé en question 2 ne contient que des valeurs plus petites que 120.

Parmi les majeur·e·s français·es, celles/ceux qui ont été condamné·e·s pour violence ont un score moyen plus élevé à ce questionnaire que l'ensemble de la population.
☐Vraie ☐Fausse ☒ On ne peut rien affirmer
En effet, les intervalles des questions 1 et 2 se chevauchent, donc ça ne permet pas de déterminer laquelle des deux moyennes est la plus grande.

Exercice 4 : Effet de réseaux sociaux sur des personne souffrant de troubles mentaux

Mme Andrieu, Mr Bonnet, Mr Courtois, et Mme Dumont sont psychologues, associé·e·s au sein du même cabinet.

Mr Bonnet, Mr Courtois, et Mme Dumont ont déjà interdit à plusieurs de leurs patient·e·s l'usage d'un même réseau social, qui diffuse de courtes vidéos. Ils/elles ont en effet constaté que l'algorithme de ce réseau social proposait invariablement à ces patient·e·s des vidéos qui les ramenaient à leur maladie, les empêchant de progresser. Mme Andrieu, pour sa part, n'a jamais interdit ce réseau social à aucun·e de ses patient·e·s.

1. Lorsque deux psychologues de ce cabinet (choisi·e·s au hasard parmi les quatre qui composent le cabinet) déjeunent ensemble, quelle est la probabilité que tous les deux aient déjà interdit ce réseau social à certain·e·s de leur patient·e·s ? *Pour répondre, vous pourrez par exemple lister toutes les possibilités quand on choisit deux psychologues de ce cabinet, puis compter dans combien de ces possibilités les deux psychologues ont tous les deux déjà interdit ce réseau social à certain·e·s de leur patient·e·s.*

On obtient la liste de cas suivante : AB, AC, AD, BC, BD et CD (où AB, par exemple, désigne Mme Andrieu et Mr Bonnet).
Parmi ces cas, seuls les trois derniers correspondent à deux psychologues ayant interdit ce réseau social à des patient·e·s.
La probabilité recherchée est donc $\frac{3}{6} = 0{,}5$.

Se rendant compte qu'au sein de leur cabinet, trois associé·e·s sur les quatre ont été confronté·e·s à ce problème, ces psychologues se demandent si cela traduit des spécificités sociologiques de la patientèle de leur cabinet, ou si c'est au contraire un phénomène général. Pour le déterminer, ils/elles décident d'enquêter auprès de 100 autres psychologues choisi·e·s au hasard (pour simplifier les calculs, le tirage au sort est effectué avec remise).

2. Dans cette question, on calcule des probabilités sous l'hypothèse que 10% de l'ensemble des psychologues aient déjà été amené·e·s à interdire ce réseau social à un·e (ou plusieurs) de leur patient·e·s.

   On considère un échantillon aléatoire de 100 psychologues, et on désigne par X le nombre de psychologue (parmi les 100) qui ont déjà interdit ce réseau social à un·e (ou plusieurs) de leur patient·e·s.

   (a) Quelle est la loi de la variable X ?

   On répète 100 fois l'expérience "choisir au hasard un·e psychologue et constater s'il/elle a déjà interdit ce réseau social", avec à chaque fois le même probabilité 10%. Donc on obtient une loi binomiale :
   $X \sim \mathcal{B}(100\,;\,0{,}1)$.

   (b) Combien vaut la probabilité $\mathbb{P}[X \geqslant 20]$.

   Pour cette loi binomiale, la calculatrice donne $\mathbb{P}[X \leqslant 19] \simeq 0{,}998\,021$.
   Donc $\mathbb{P}[X \geqslant 20] = 1 - \mathbb{P}[X \leqslant 19] \simeq 1 - 0{,}998\,021 = 0{,}001\,979$.

3. Quand on réalise l'expérience, on obtient que 45 psychologues sur les 100 ont déjà interdit ce réseau social à un·e (ou plusieurs) de leur patient·e·s. Votre réponse à la question précédente conforte-t-elle l'hypothèse (utilisée à la question précédente) selon laquelle 10% de l'ensemble des psychologues aient déjà été amené·e·s à interdire ce réseau social à un·e (ou plusieurs) de leur patient·e·s ?

Ce qu'on a observé est un cas particulier de l'évènement $X \geqslant 20$ dont on vient de calculer la probabilité. Cette probabilité était très faible, donc l'évènement $X \geqslant 20$ aurait eu très peu de chances de survenir si X suivait bien la loi $\mathcal{B}(100\,;\,0{,}1)$. Si cela a eu lieu, c'est vraisemblablement qu'en fait X ne suivait pas la loi $\mathcal{B}(100\,;\,0{,}1)$, c'est à dire qu'il n'y a vraisemblablement pas 10% de l'ensemble des psychologues qui ont déjà été amené·e·s à interdire ce réseau social à un·e (ou plusieurs) de leur patient·e·s.